Ken Buetow, Ph.D.
2008 caBIG™ Annual Meeting
Tuesday, June 24th (8:30 am)
Omni Shoreham Hotel, Washington, D.C.

Pathway to a New Model for Biomedicine: caBIG™ and Beyond

Today's focus is trying to explore and further describe what it really means to bring this next generation biomedical agenda of blending information and technology and new paradigms together in a unified whole, and how we will be able to enable a whole new generation of biomedical research and biomedical paradigms through the activities that this audience is doing.

What I hope you will all do is help me in welcoming our fellow panelists this morning who will discuss their own experiences with caBIG™, as well as their visions of how we will move into the future of biomedicine. They'll describe, in particular, their enabling and pioneering work in personalized medicine.

So at my extreme left we have Dr. Peter Traber. Dr. Traber is the President and CEO of Baylor College of Medicine, and we're honored that he's traveled to be with us from Houston. Dr. Traber is a respected academic leader and physician/scientist with more than 20 years of experience in the healthcare industry. At Baylor, he is currently working to build a new model of patient care, from the ground up, for the delivery of personalized medicine.

To Dr. Traber's right, we have Dr. Lou Weiner who's the Director of the Lombardi Comprehensive Cancer Center at Georgetown University, just around the corner here in Washington DC. He's recognized internationally for his work in both laboratory and clinical research, where he's focused on new therapies that utilize a patient's own immune system. Dr. Weiner is a colleague and a friend. We go back a long time to my pre-NCI days, back in the Fox Chase Cancer Center when both of us were pioneering this new and exciting space. I'm personally delighted that he and his organizations have embraced caBIG™. Dr. Traber and Dr. Weiner are both at the forefront of the personalized medicine revolution, and they'll be sharing their insights and perspectives from the front line.

Then, to Lou's right is Dr. Mark Adams—a friend and actually a familiar face to most of you in the caBIG™ program. He's the caBIG™ program manager from Booz Allen Hamilton. Dr. Adams has more than 15 years of experience in the field of bioinformatics with a focus on designing and implementing high throughput molecular biology and bioinformatics systems. Mark is one of the rare birds of our field who understands both biology and informatics, and he also participated in the first generation of molecular medicine companies. He's going to be describing one of the finest achievements of caBIG™ and one of our research poster children: the use of caBIG™ to enable the data portal of The Cancer Genome Atlas.

Before we kick off the panel, I'm going to start the morning with a few framing remarks that hopefully set the context for this day's class of activities by identifying where we sit today in the caBIG™ program and, more importantly, the trajectory that we see it undertaking over the next period of time so that we can all synchronize our watches and figure out how we're going to get to our next destination.

What I'm going to be talking about in my presentation today is what we would call the pathway to a new model for biomedicine, caBIG™ and beyond. To infinity and beyond. I'm going to talk about how our current capabilities in caBIG™ are enabling a new brand of clinical and biomedical research with a transition into how we see caBIG™ helping to underpin the paradigm that we're going to hear our speakers address. How are we going to leverage what we're doing in clinical research into this next generation biomedical enterprise where we actually create a unified health system rather than a collection of disconnected sectors? We believe that you, as the caBIG™ community, are going to be essential to performing that task.

I like to start out sometimes with a literary context. I think anyone who's engaged in biomedicine today recognizes that we live in a sort of "Dickinsonian universe" at times. It is unquestionably the best of times, in many ways. There is no question from anyone that the twenty-first century is the biomedical century. Just the same as the twentieth century, perhaps, was the century of physics and chemistry.

The twenty-first century is the century of the biomedical revolution, and the transformation, we hope, of the healthcare enterprise. We've made tremendous progress in understanding the underlying molecular basis of disease over the last few years. In fact, if you extend this thought, we've gone only 10 percent of the way there. Given that we've only found associations with 2,000 genes, it is unlikely that we won't find some degree of disease association within the entire genome and other extra genomic gene class features.

We also live in an interesting period where we're now starting to translate these findings of genetic basis of disease. We are learning how we can individualize the basis of disease and how we can have more effective therapeutic treatments by taking into considerations either individual variation within a specific disease or individual variation in the person themselves and how they will respond to therapeutic agent based on their constitution or other molecular components.

There is an explosion of the number of diagnostic and therapeutic class characteristics and diagnostics that we can perform in order to test these types of activities. However, in true Dickinsonian nature, we're also living in the worst of times. Many of us in biomedical research recognize that the resources that we're facing have never been more limited. Moreover, just the basic demographics of the entire country are driving us to the point where we're going to shortly be in crisis. We are facing the aging baby-boomer explosion in the load on healthcare. We recognize that cancer, very shortly, will be the number one cause of death. And we recognize that if we look at just some of the simple predictions of how much we're spending on healthcare delivery, we're growing at an unsustainable rate. You can see—projecting into the future as the baby boomers start to actually hit the scales—that we're going to spend maybe as much as 20 percent of our economic capacity on just providing basic healthcare.

One of the reasons we're facing this is that we have fundamental disconnects between all of the components of the biomedical enterprise, as well as overwhelming amounts of data and challenges in communicating that data meaningfully from sector, to sector, to sector. We find ourselves having to constantly reinvent the wheel, and what we always find as well is that we never seem to have enough money to do anything right, but we always seem to have enough money to do it over. So, what we do is create and proliferate silos of information in basic research, clinical translational research, and healthcare delivery. We need to actually figure out ways that we can leverage the insights, infrastructure, and communities associated with each of these things in novel ways.

To that end, we think that the Biomedical Informatics Grid, the caBIG™ activity, is an essential component both to transform and to support this new research paradigm—this molecularly-based medicine paradigm where we have the capacity to tailor treatments to individual disease and individual characteristics as well as then help underpin this next generation activity in biomedicine.

From my perspective, one of the tremendous accomplishments we made over the last five years of caBIG™ activity is the connecting and generation of community. We have unquestionably created a rich tapestry of interconnected individuals, organizations, and institutions across the domains of biomedical research. We actually have, in this room alone, individuals representing clinical paradigms, basic research paradigms, and different institutions. We have individuals representing different types of analytic approaches and types of information that are all being integrated into a rich tapestry that we call the caBIG™ network or the caBIG™ family.

But the key to this in caBIG™ has been the "warp and woof," the tapestry, the basic threads that underpin this. So, as important as the community is, I would say, by far the most important contribution of this community to the broader biomedicine is the stitches—the fabric on which we can build a new generation biomedical paradigm.

We hear a lot of discussion about what caBIG™ is all about. At the end of the day, caBIG™ is all about standards, interoperability, data sharing, and connectivity. What we're delivering through this caBIG™ infrastructure is the capacity to bring all of these individual components together by the use of international standards and by defining how we interoperate, both with applications and infrastructure that we build within the caBIG™ program as well as with commercial opportunities. We are working to share valuable and invaluable contributions by different groups and connect those through the next generation World Wide Web infrastructure we call caGrid.

caBIG™, as of today, actually is a very rich and well-annotated tapestry. We have more than 300 software applications and more than 40 end-user applications. It's a truly incredible universe of capabilities that are now present as part of the caBIG™ framework. Moreover we're moving to make these much more accessible by a broader community through bundling them in framework such as the clinical trials framework, life sciences distribution, and—even for the non-technical framework—the data-sharing and security framework.

We've launched a next generation connectivity infrastructure through caGrid in our services-oriented architecture, and lastly, we're moving aggressively forward into providing the support necessary to sustain this larger scale activity with our Knowledge Centers, as well as moving forward to bring the commercial sector into the caBIG™ family in licensed service providers.

As we stand today, we are poised to take a bold step into this next generation of biomedicine with a collection of standing infrastructure within the NCI—46 NCI-designated Cancer Centers actively deploying caBIG™ and 10 NCI Community Cancer Centers—blending our delivery system with our research system. Of equal importance, caBIG™ is also now moving into the broader federal health architecture in the efforts of the Office of the National Coordinator as part of the National Health Information Network. caBIG™ components are now being actively recycled to facilitate the federal health architecture and to be part of the federal health gateway.

caBIG™ is now also moving globally with active partners, such as our NCRI colleagues from the UK. We also now have active participation from the People's Republic of China in both Beijing and Shanghai, and we are forging a new partnership with our colleagues in India, as well as exploring relationships with colleagues in Latin America.

So, we're well on our way within the research context of the goal of patient-centric, targeted care in cancer and what we hope will be a model for other diseases. We're on the path, at least at an IT level, to connect these individual pieces so we can deliver on the personalized medicine goal.

However, I guess I want to highlight that we sit at a complicated transition as well. While we have this technical capability to bring together the individual components of biomedical research, it's not *a priori* clear (given the challenges that I just articulated in terms of resources, disconnected communities, and data silos) that we can actually—from where we sit today in our biomedical enterprise—get to the destination that we long to achieve. It's in the notion that quite commonly in bio—in any space, but clearly where we sit in biomedicine today—that sometimes the problems that we've created actually can't be solved by the same approaches that we've used to get here today.

I'd like to spend the last part of my talk discussing how we move forward in this agenda. How we see taking the caBIG™ inside viewpoint and leveraging this to help address this next generation biomedical paradigm, and how we can build on the shoulders of what you all are doing and have you all carry biomedicine into this next generation activity.

So what's our concept? Well the idea here is to blend discovery, clinical research, and clinical care so that there's much more of a continuum as opposed to these isolated separated silos. The goal of this is having faster clinical validation; faster, more efficient patient recruitment into trials; and improved clinical trials outcomes due to improved patient selection. But also on the flip side of this, on the care delivery side, the goal is a more robust, evidence-based, and faster adoption of the findings of the biomedical

space into the care delivery system and overall a reduced cost of infrastructure from actually leveraging each other's investments and leveraging each other's insights, knowledge, and capacity as we move forward.

So, you all say that's so radical. How could we possibly do that? I'd like to point out that, in fact, this experiment has been going on for a long time, and those of you in the cancer community should recognize that this is not a heretical idea, especially if we look in pediatric cancer as an example. In many ways for the last 30 years, it's been the paradigm of what personalized medicine could/should look like; it was the first adopter of genome-wide characterizations. We, now-a-days "poo poo" the notion of what we were doing at that point of time, but very early on in the diagnosis of childhood cancer, in particular in leukemia's and lymphomas, we used the first generation of genome-wide scans. We called them karyotypes. They provided a full characterization of the genome constitution with the technologies that existed at the time and allowed us to distinguish—what under a microscope would have been absolutely identical appearing cells—and identify that these were very different underlying diseases based on their genetic constitution.

What we also found is that those different molecular characterizations made a difference as to how you treated people. We saw that leukemia in this instance wasn't all one disease as it appeared under a microscope, but in fact was 15 to 20 different diseases based on the genetic constitution of those diseases. More importantly, what we found is that those different diseases actually should be treated in different ways if we wanted to have effective outcomes. We actually have a very successful paradigm of how we translate molecular medicine into ongoing care.

This translation unquestionably has had tremendous benefits. A little over 30 years ago, identification of childhood cancer, especially leukemia, was a death sentence. Now, through these next generation approaches, there are tremendous survival rates, and for certain forms of cancer we are approaching 95–98 percent survival rates given the appropriate designations and characterization of the disease. What, in part, drove that?

One of the things that drove that is even though, thankfully cancer in children is a relatively rare disease, what we actually recognized is that there was, from the get go, a blending of clinical care and clinical research in the context of performing childhood cancer treatment. What we see is that a phenomenally large proportion of children participate in clinical trials. More than 63 percent of children who are seen at a pediatric center are actually enrolled and are active participants in a clinical trial. This is unprecedented if we look at what we see in the context of adult cancer where those numbers are probably less than 5 percent.

What does this mean? The notion that children are participating in this clinical trials framework means that researchers and practitioners are able to, on a real time basis, correlate experimental laboratory data with clinical data. They have full access to treatment history, pathology, and outcomes, and biomarker discoveries are applied to subgroup patients for experimental treatments as rapidly as the information can be compiled. Clinical data are utilized continuously to evaluate outcomes, and researchers evaluate and define evidence-based strategies in the same context that the care is being delivered. Then, because those environments are combined, care providers have immediate access to the state of the art findings that are occurring from clinical research.

Of course, information flow is critical to the success of this activity. You can't do this if all of the information is segregated and siloed. It has to be interconnected. We believe the caBIG™ framework provides a way that we can actually bring together these fragmented segments of biomedicine, the healthcare delivery system, the clinical research environment, as well as the regulatory reporting environment. We do this by taking advantage of it and synthesizing this very complex landscape of standards and working together with both communities—again in a partnership between HL7, CDISC, and an active participant, the NCI—to create common information models and common information standards. For instance, the BRIDG biomedical research domain group has created a universal lingo that we can use to talk across these environments.

Because of this, then, we conceptually have the standards that allow us to translate information between these individual domains. Moreover, caBIG™ has been developing infrastructure that allows this connectivity. For instance, in the healthcare delivery system we have tools that facilitate the exchange of these HL7 messages into resources that can be either leveraged to study outcomes in an individual hospital or to be shared into the clinical research enterprise. We also have built rich tools that facilitate this clinical research enterprise by substantiating the standards so that the information can float seamlessly into this framework.

We've built our infrastructure so that the pipeline of information from clinical trials can flow electronically and seamlessly to regulatory reporting, both within the NCI, but more importantly to our pharmaceutical partners and to the FDA. We have in place a collection of research resources that can be leveraged in this broader, integrated framework that supports the clinical encounter, and supports the availability of *in vivo* imaging. That support, the use and reuse of pathology, supports the new molecular medicine paradigm.

For instance, we have a tool that will track who's enrolled in what types of activities in a clinical setting, and we can share this information with broader data warehouses, with appropriate consent to do so. We have the ability to share—both within a research setting and in a clinical setting—the rich imagery that's part of the modern cancer enterprise (and expanding in other domains as well) as the definitive means by which diagnosis is done and quite commonly the definitive means by which people are charting progress of cancer therapeutics.

We have the ability to share biospecimen resources across the broader community with a collection of tools in a research setting that can also then bridge back to the care setting so we don't have to have biospecimens be the rate-limiting step of any of the work that we're attempting to do in biomedical research. And lastly, we have a very rich infrastructure that supports the molecular paradigm of medicine and allows us to go and actually not only conduct next generation personalized medicine research but also support the execution of personalized molecular medicine in a clinical and care delivery setting.

What we see by blending these two universes is that we can share a common infrastructure—reducing the cost of both universes—and build a bridge that will facilitate the molecular medicine delivery back to the clinic. In this instance, by having this infrastructure in place, we have the capacity to both support next generation clinical research and also, more importantly, inform next generation practice outcomes and bring molecular medicine into the clinical practice environment.

However, if we're going to do this on a broader scale, we need to bring more participants to the table. Not to take anything away from our exciting caBIG™ community that we have today and the focus on clinical research, but we need more players. These include both our clinical communities, as we currently have, but also a much more meaningful engagement of care providers and consumers. We need to have a full partnership of funders—not just government as has been the case to date with caBIG™—by bringing in other players to help underpin this infrastructure.

We need to meaningfully bring in the life scientists so that their insights are not foreign to the broader enterprise, and we need to blend and connect both the research infrastructure and the electronic health infrastructure through next generation health information technology and the partnership of large-scale information technology providers.

What does this actually look like? We need to broaden our fabric, bringing new members and new capabilities, technical partners, pharmaceutical industry, personalized medicine, technology companies, diagnostic companies, and a variety of other groups into this much broader fabric. But once we do this, then, it allows us to truly address this next paradigm. So, what does it actually look like?

From the perspective of a researcher, what we can do now is conduct next generation research by having this full connection across the continuity of the whole healthcare delivery system from delivery all the way to research, so we can facilitate for our research activities and participant selection. We have access to

biospecimen collection. It's not *de novo* collected for a specific study or owned by an individual researcher, but is community owned.

We have the ability, then, with rich data resources on tap to perform correlations not necessarily from the work I do in my laboratory, but leveraging the accumulating collections of observations that are put against individual patient samples, individual participant work, as well as large collections of clinical observations. Then, we can drive the research agenda by lowering the barrier necessary to connect and conduct these next generation personalized trials by recycling information routinely rather than having to *de novo* collect it each time we conduct a research investigation.

However, it's not just a benefit to the researcher. We believe that all people have to find something in this for them to make this a workable parameter, a workable agenda. So, we also see that in fact, when there are opportunities for the clinical practice arena through the use of electronic health records, then we will have the ability to track outcomes information. Accumulating these data and providing real time access to individual hospital or practice information will broaden the participation of individual patients, letting them travel to and connect with any physician or any community hospital that can deliver the flavor of care that they're interested in receiving as well as broadening participation for the physician.

From the patient perspective we see this, in many ways, as the thin edge of the wedge—the true disruptive piece of this whole paradigm. By empowering patients, they then have access to their genomic and other clinical information in a manner unprecedented today. They can use this information prospectively to determine what types of interventions or what types of prevention strategies they choose to undertake, but they also have the power to actually direct this information across the paradigm and share it with care providers wherever they so choose to share this.

What we see is this rich new tapestry empowered by the underlying caBIG™ infrastructure that permits us then to connect the cancer research community and then becomes a model for other disease communities where infrastructure that links separate silos of healthcare information actually enables this new paradigm. So, caBIG™, as we can see, is already helping us join hands, not only within the United States, but across the globe. What we're seeing is as caBIG™ becomes BIG-inside, we have the capacity to explore this new approach to consumer-driven healthcare where we actually have the patient-centric approach that you'll hear future speakers talk about.

What we're hoping is that "Beyond caBIG™" enables this next generation of personalized medicine that is preemptive, predictive, and participatory. So just perchance that you're worried that we're too far out ahead, I can tell you that our leader at the National Cancer Institute, Dr. John Niederhuber, has not only not discouraged us from this bold class of thinking but also has actually mandated that our next generation thinking needs to be moving not only on delivering on the promise that we have in terms of our clinical research environment but also actually challenging each and every one in this audience to be imagining the destination we want to get to and to see to it that what we are building facilities to get us to that destination.

We are quite encouraged and excited about the potential of moving beyond our core cancer research agenda to exploring this new blended healthcare and research environment with a caBIG™ and beyond approach that really allows us to usher in the age of molecular medicine.

[End of Recording]